

RASSCLE II Database Design Model

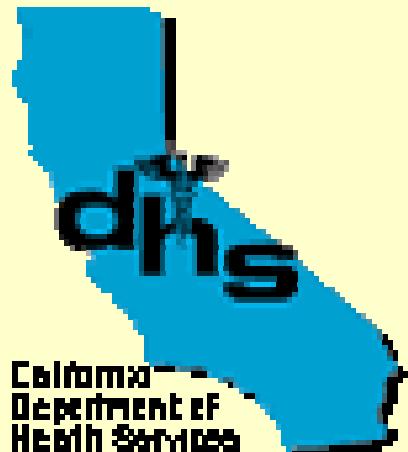
PHIN Compliance

David Goldsworthy
Abdul-Malik Shakir

California Department of Health Services

May 2005

Presenters



- **David Goldsworthy**

Lead Technician for the California Department of Health Services (CA-DHS) Childhood Lead Poisoning Prevention Bureau (CLPPB) Response and Surveillance System for Childhood Lead Exposure (RASSCLE) project.

- **AbdulMalik Shakir**

Principal Consultant with Shakir Consulting and Senior Advisor / Lead Data Architect for the California Department of Health Services Childhood Lead Poisoning Prevention Bureau Response and Surveillance System for Childhood Lead Exposure project.



Presentation Background

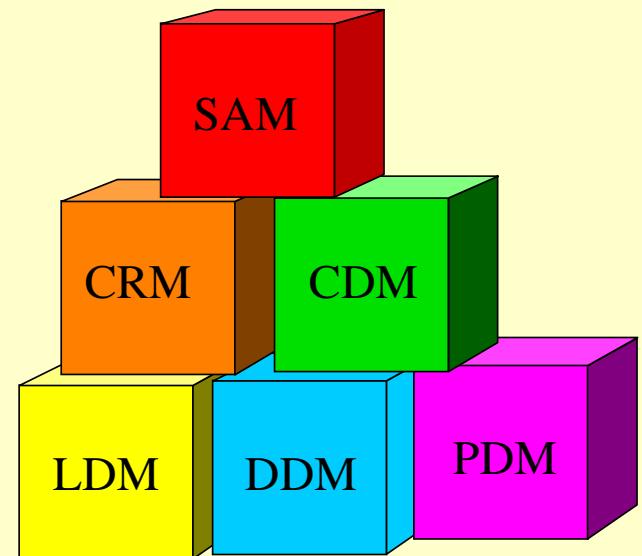
- The Response and Surveillance System for Childhood Lead Exposure (RASSCLE) is used by CLPPB to collect information on children found to have elevated blood lead levels.
- The current RASSCLE system is being re-engineered as a state-wide web-based information system.
- As part of the redesign it was necessary to construct a database design model for use in implementing a relational database structure for the RASSCLE II system.
- CLPPB used the PHIN Logical Data Model (PHLDM) to guide their design of the RASSCLE II database.

Presentation Overview

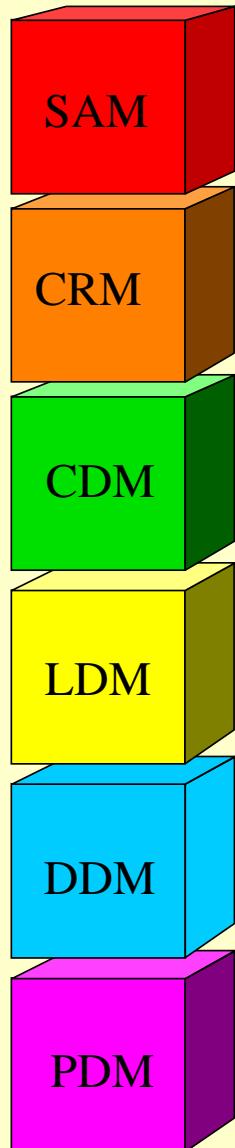
- This presentation will provide an overview of the process used to derive the RASSCLE II Database Design Model.
- It will demonstrate how the PHLDL and HL7 RIM were used to create a RASSCLE-specific logical data model.
- The parameters used by the RASSCLE II team as a definition of PHLDL compliance will be shared.
- The design trade-offs that resulted in the RIIDDM will be explored along with the criteria used to arrive at the particular design employed.

Hierarchy of Data Model Types

- Subject Area Model
- Class Relationship Model
- Conceptual Data Model
- Logical Data Model
- Database Design Model
- Physical Database Model

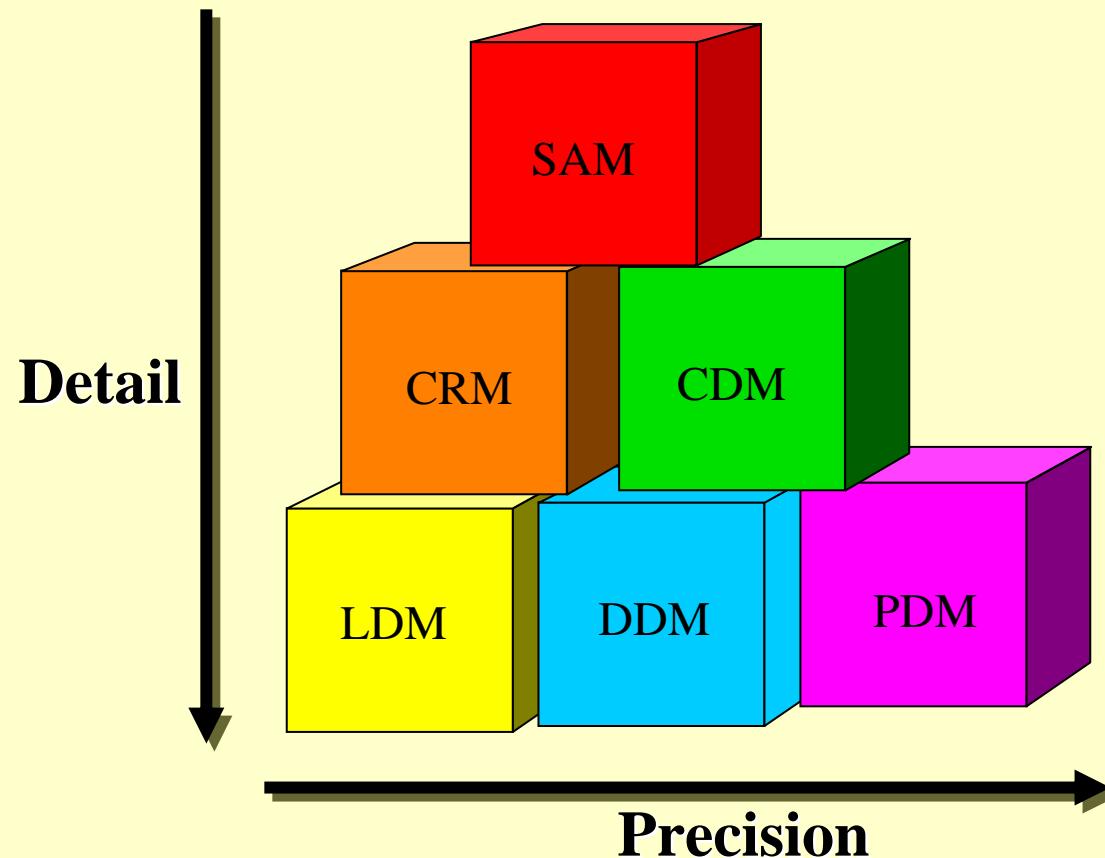


Hierarchy of Data Model Types



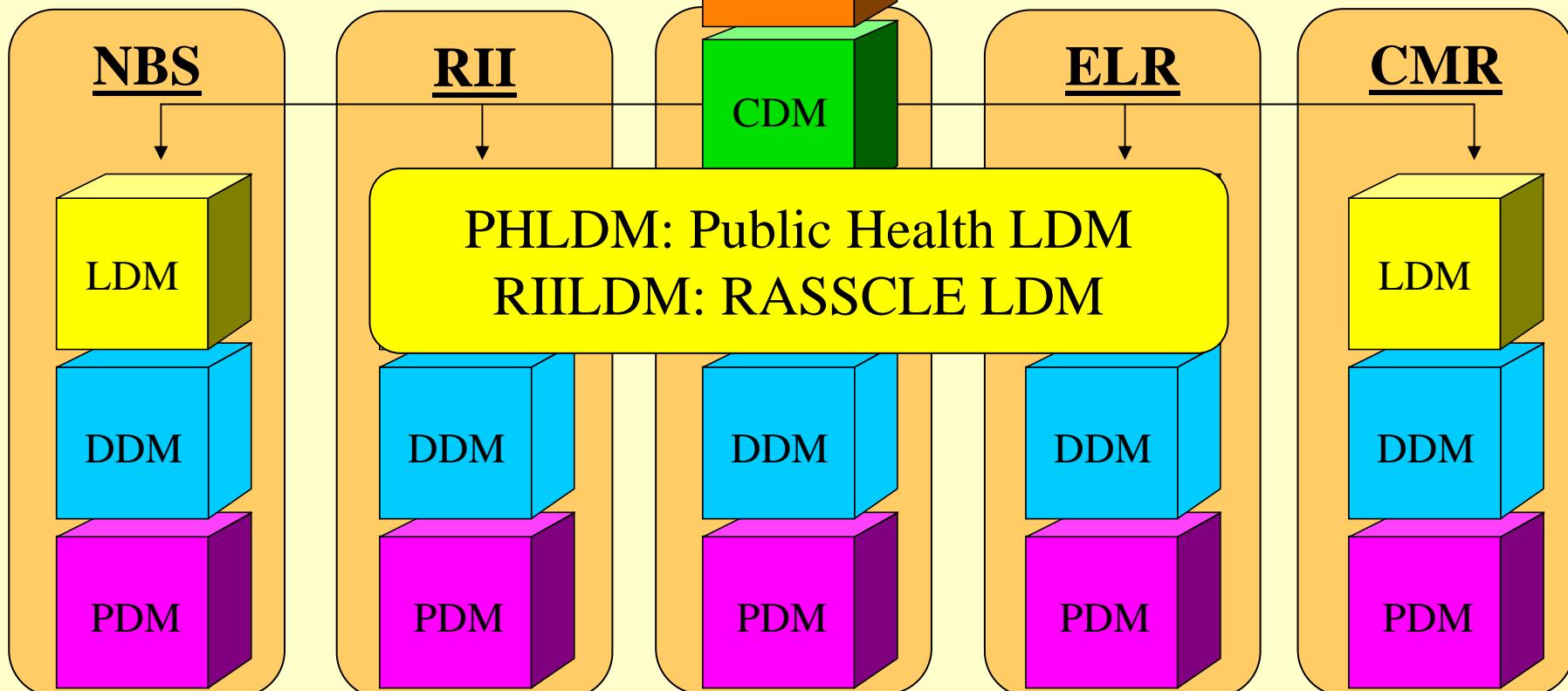
- **Subject Area Model**
 - Contains subject areas and connections only
 - Used for high level planning and project scoping
- **Classes Relationship Model**
 - Contains subject areas, classes, and associations only
 - Used for high level analysis and project estimating
- **Conceptual Data Model**
 - Contains subject areas, classes, attributes, and associations
 - Results from detail level analysis and is often a project deliverable
- **Logical Data Model**
 - Contains subject areas, classes, atomic attributes, associations, and primary keys
 - The completion of the most detailed analysis and the beginning of database design
- **Database Design Model**
 - Contains table spaces, tables, columns, datatypes, primary and foreign keys
 - The completion of the most detailed design and the beginning of database construction
- **Physical Database Model**
 - Contains data definition language (DDL) required to create tables, indexes, and constraints
 - Database construction, the most detailed, DBMS specific data model expression

Hierarchy of Data Model Types

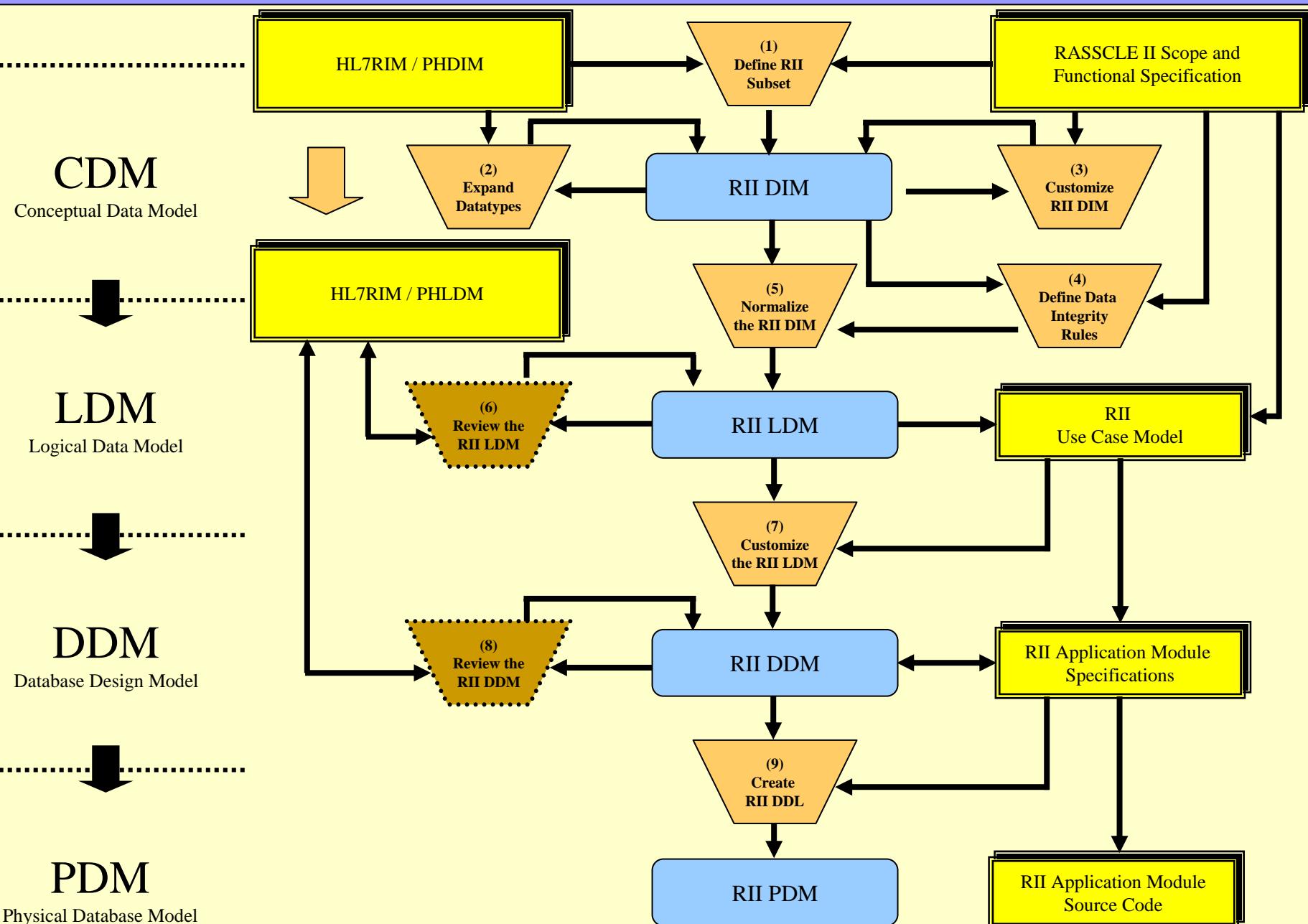


Multiple PDMs from one CDM

HL7RIM: Health Level Seven Reference Information Model
PHDIM: Public Health Domain Information Model



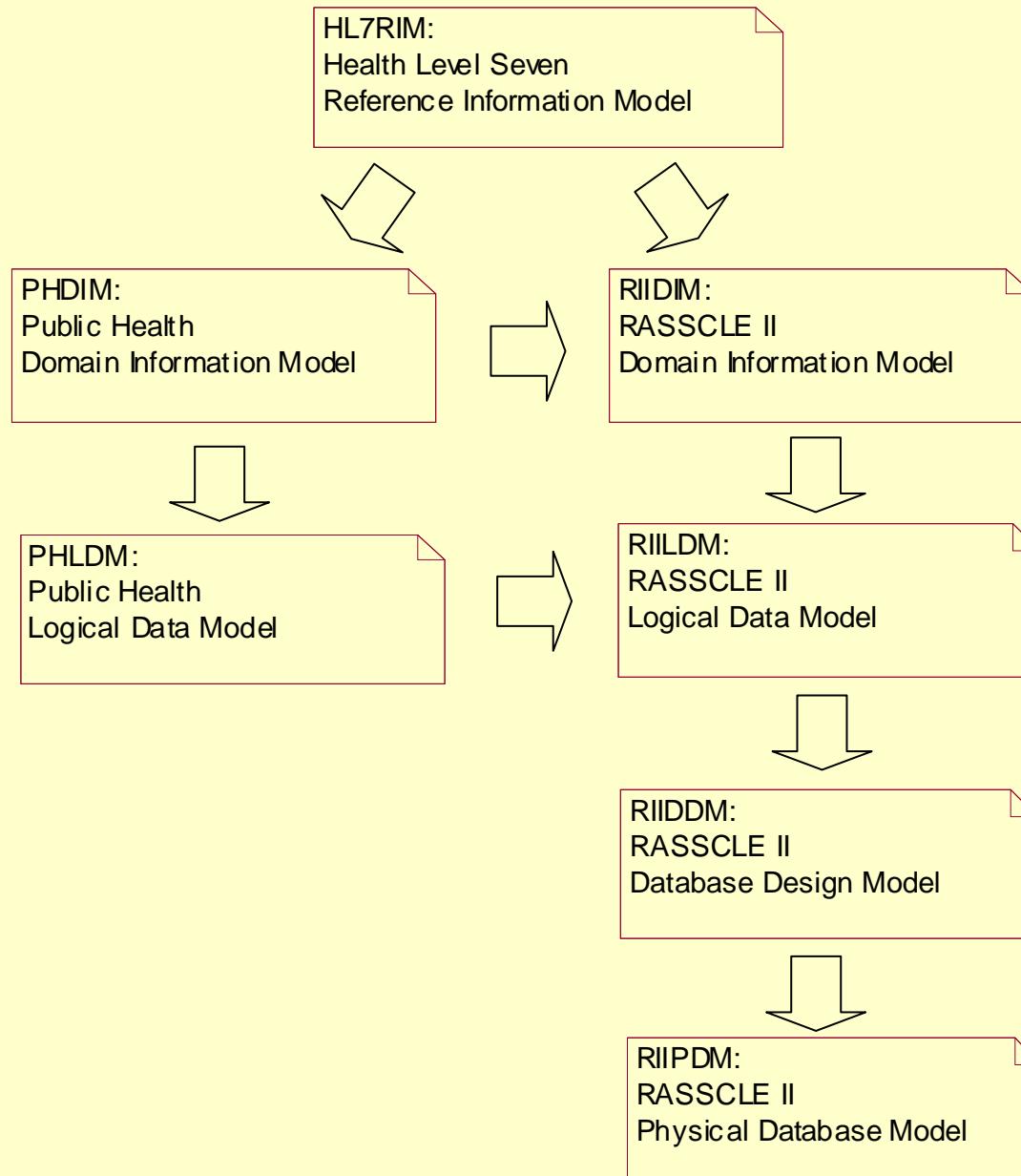
From the PHDIM to the RIIPDM



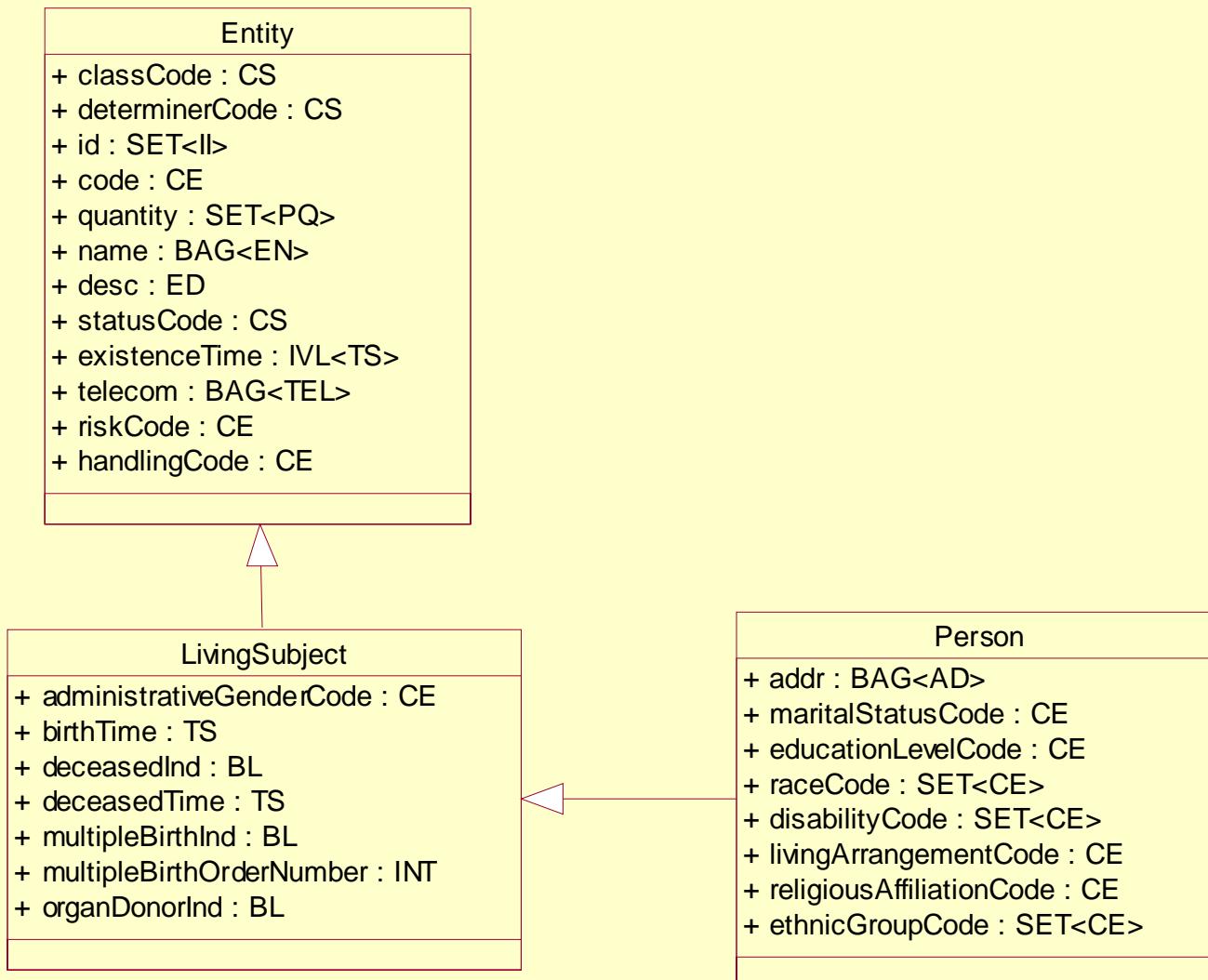
Relevant Information and Data Models

- **HL7RIM:** HL7 Reference Information Model
- **PHDIM:** Public Health Domain Information Model
- **RIIDIM:** RASSCLE II Domain Information Model
- **PHLDM:** Public Health Logical Data Model
- **RIILDM:** RASSCLE II Logical Data Model
- **RIIDDM:** RASSCLE II Database Design Model
- **RIIPDM:** RASSCLE II Physical Database Model

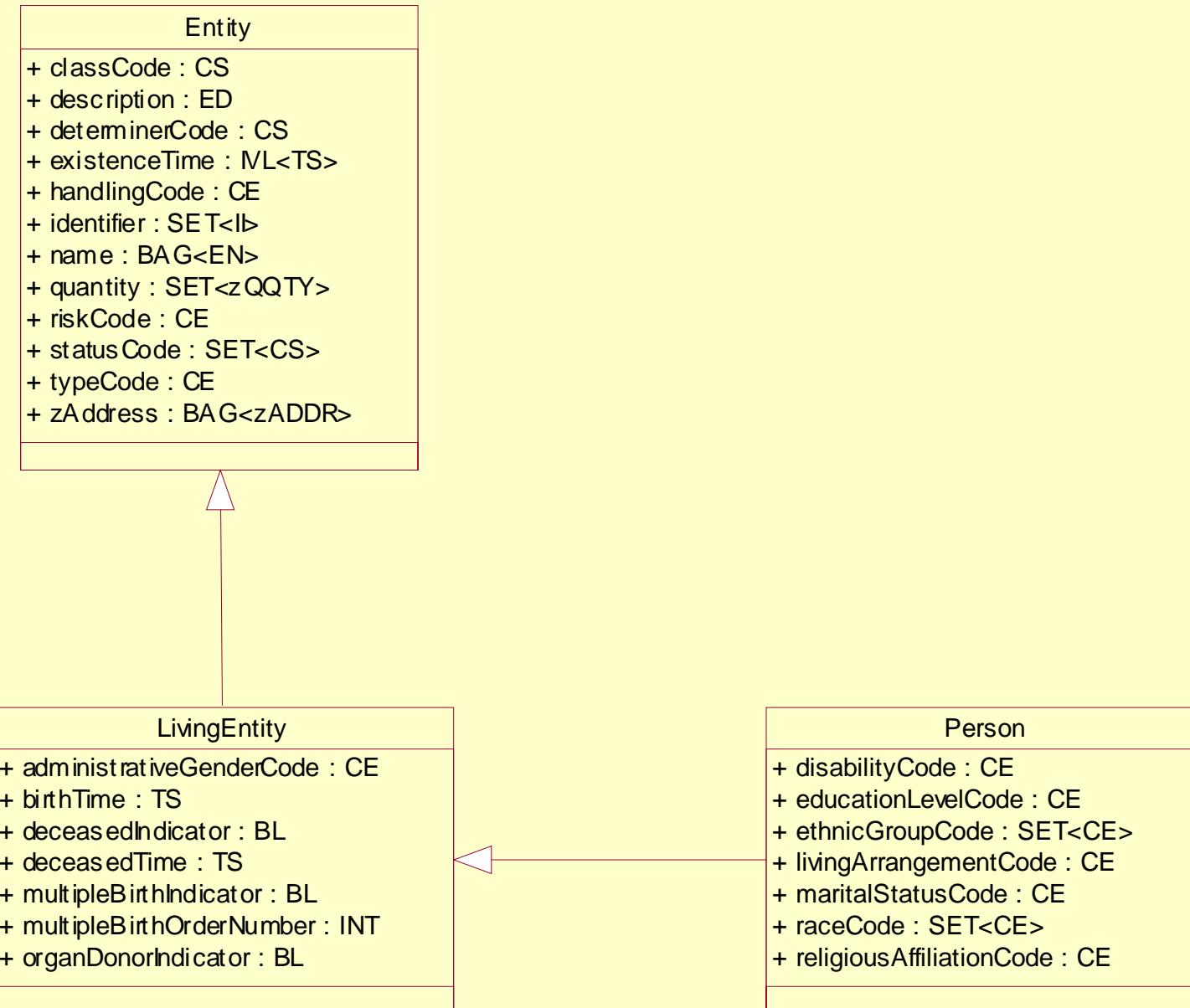
Inter-Model Derivation Relationships



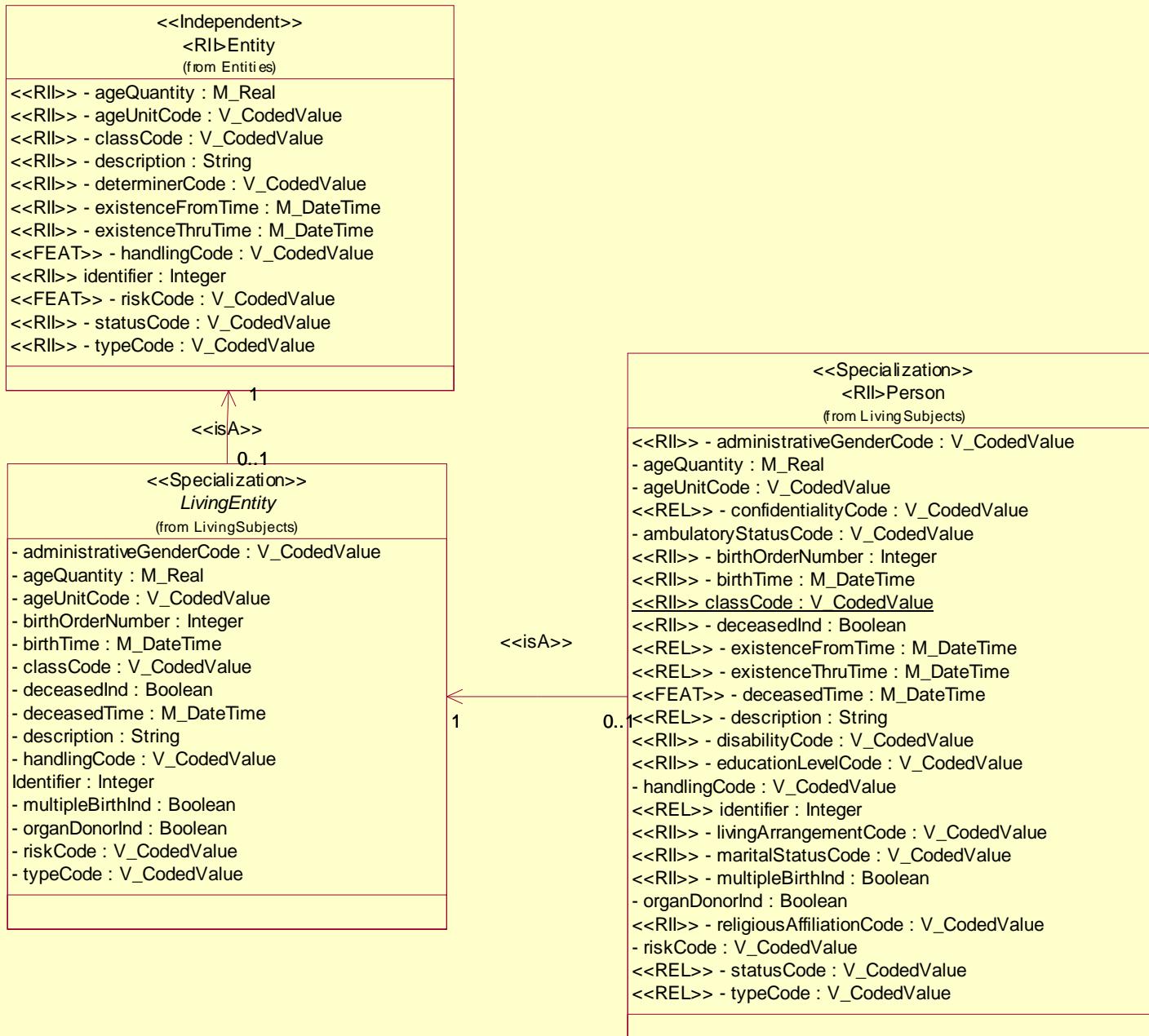
HL7 Reference Information Model



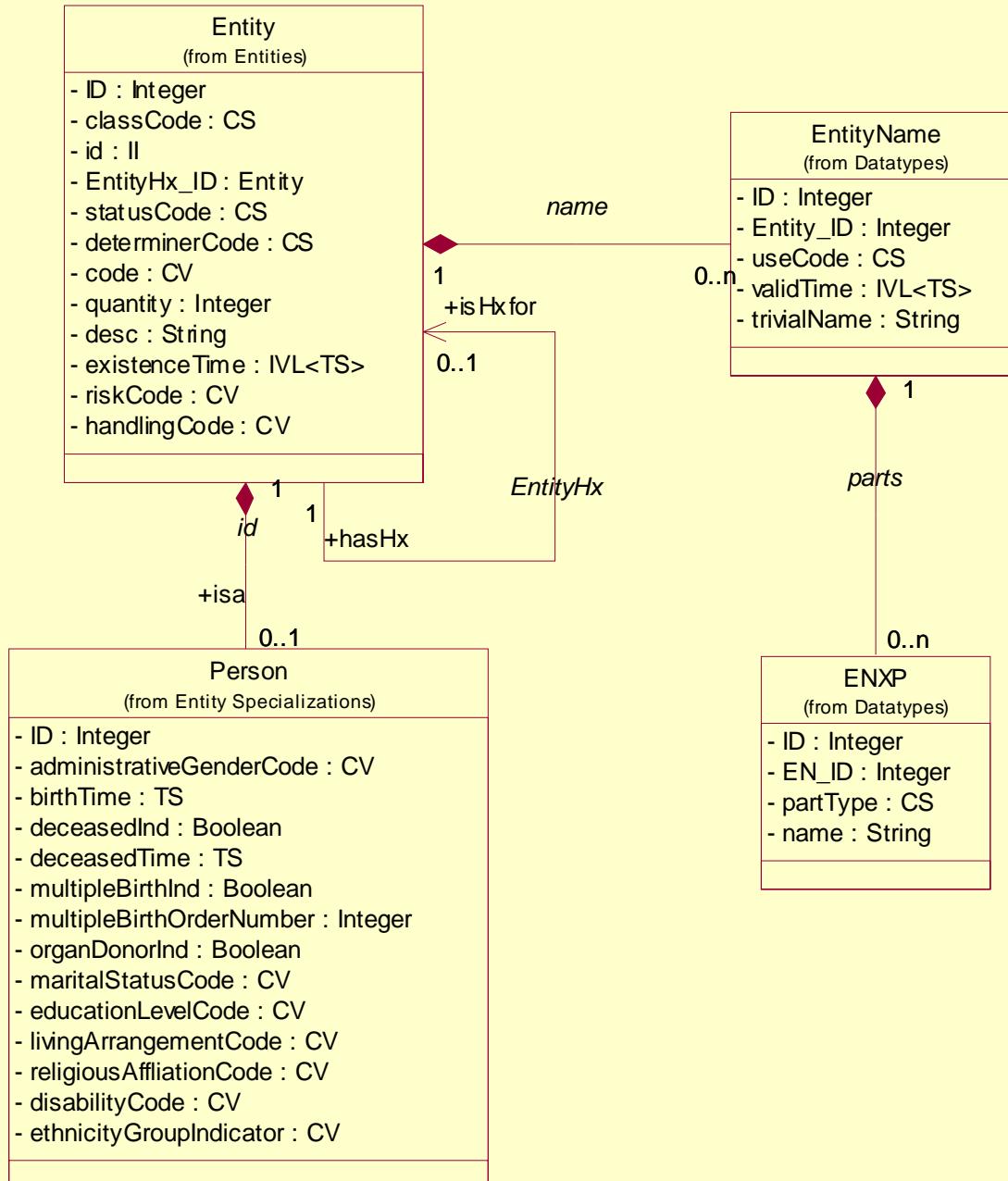
Public Health Domain Information Model



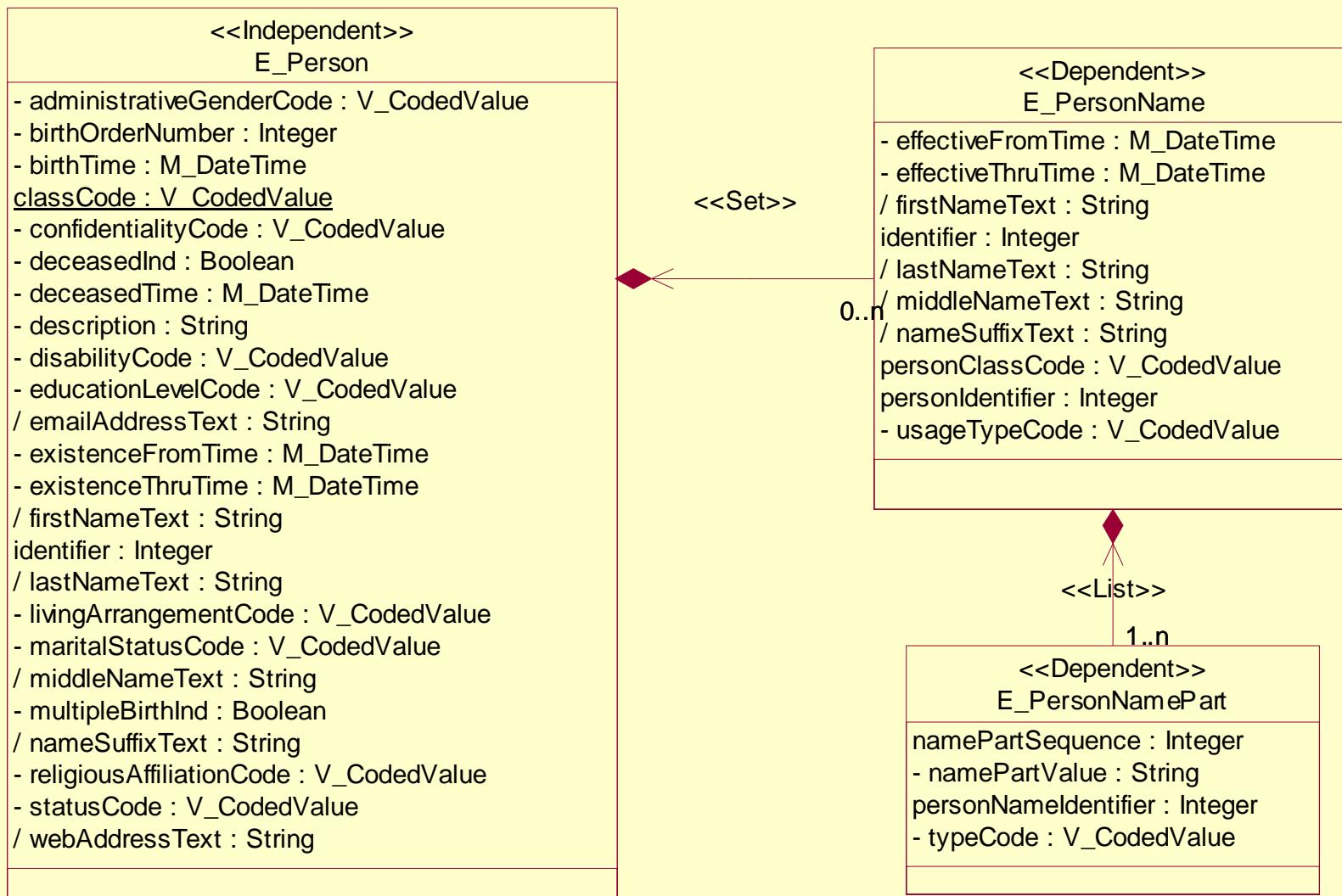
RASSCLE II Domain Information Model



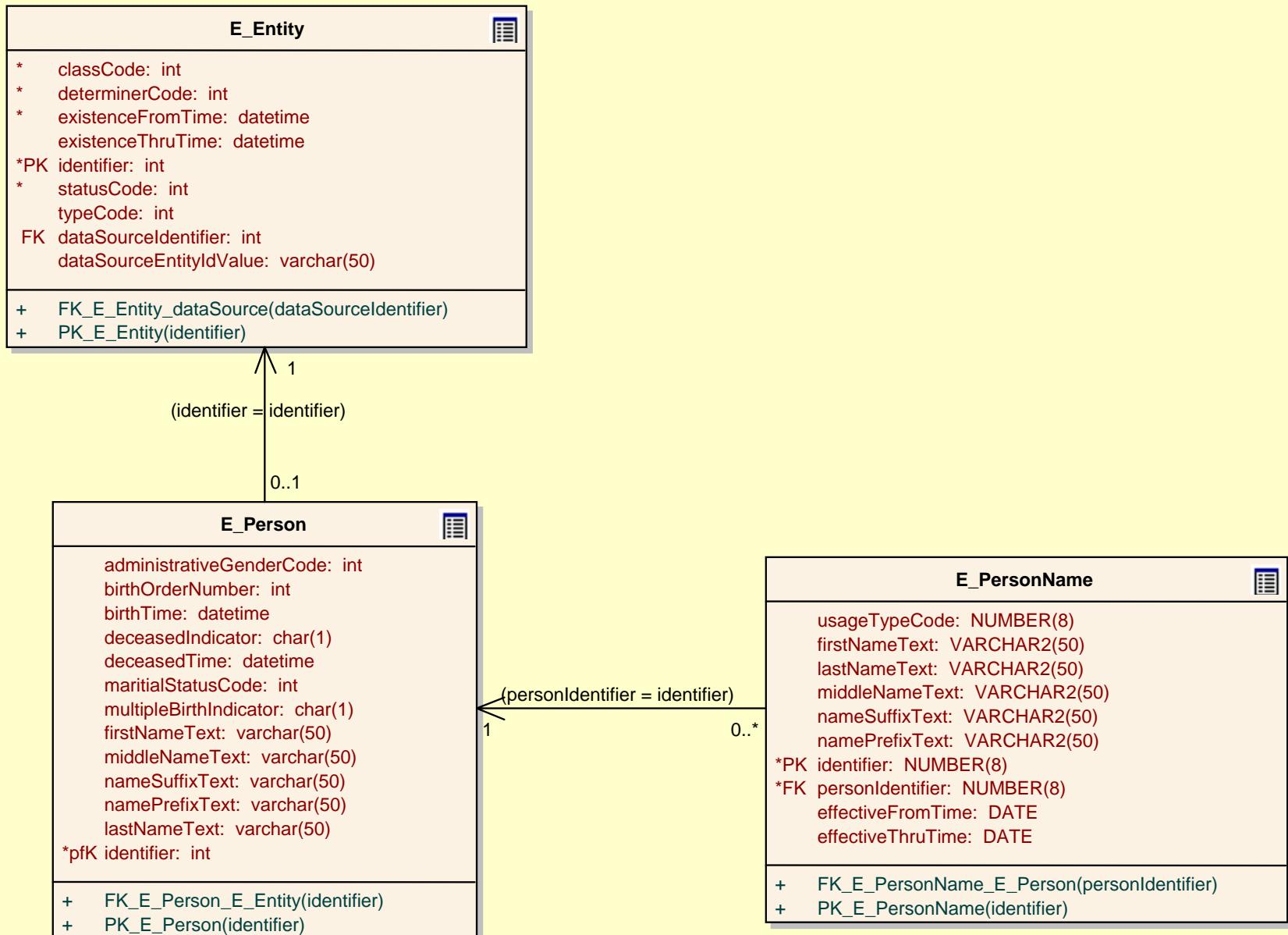
Public Health Logical Data Model



RASSCLE II Logical Data Model



RASSCLE II Database Design Model



RASSCLE II Physical Database Model

```
• CREATE TABLE E_Entity (
•   classCode int NOT NULL,
•   determinerCode int NOT NULL,
•   existenceFromTime datetime NOT NULL,
•   existenceThruTime datetime,
•   identifier int NOT NULL,
•   statusCode int NOT NULL,
•   typeCode int,
•   dataSourceIdentifier int,
•   dataSourceEntityIdValue varchar(50)
• )
• ;
• ALTER TABLE E_Entity ADD CONSTRAINT PK_E_Entity
• PRIMARY KEY (identifier)
• ;
• ALTER TABLE E_Entity ADD CONSTRAINT
• FK_E_Entity_dataSource
• FOREIGN KEY (dataSourceIdentifier) REFERENCES
• Z_DataSource (identifier)
• ;
• CREATE TABLE E_Person (
•   administrativeGenderCode int,
•   birthOrderNumber int,
•   birthTime datetime,
•   deceasedIndicator char(1),
•   deceasedTime datetime,
•   maritalStatusCode int,
•   multipleBirthIndicator char(1),
•   firstNameText varchar(50),
•   middleNameText varchar(50),
•   nameSuffixText varchar(50),
•   namePrefixText varchar(50),
•   lastNameText varchar(50),
•   identifier int NOT NULL
• )
• ;
• ALTER TABLE E_Person ADD CONSTRAINT PK_E_Person
• PRIMARY KEY (identifier)
• ;
• ALTER TABLE E_Person ADD CONSTRAINT
• FK_E_Person_E_Entity
• FOREIGN KEY (identifier) REFERENCES E_Entity
• (identifier)
• ON DELETE CASCADE
• ;
• CREATE TABLE E_PersonName (
•   usageTypeCode NUMBER(8),
•   firstNameText VARCHAR2(50),
•   lastNameText VARCHAR2(50),
•   middleNameText VARCHAR2(50),
•   nameSuffixText VARCHAR2(50),
•   namePrefixText VARCHAR2(50),
•   identifier NUMBER(8) NOT NULL,
•   personIdentifier NUMBER(8) NOT NULL,
•   effectiveFromTime DATE,
•   effectiveThruTime DATE
• )
• ;
• ALTER TABLE E_PersonName ADD CONSTRAINT
• PK_E_PersonName
• PRIMARY KEY (identifier)
• ;
• ALTER TABLE E_PersonName ADD CONSTRAINT
• FK_E_PersonName_E_Person
• FOREIGN KEY (personIdentifier) REFERENCES E_Person
• (identifier)
• ON DELETE CASCADE
• ;
```

RIIDDM Design Constraints

1. Meet the documented RASSCLE II information and functional requirements.
2. Remain consistent with the PHIN Logical Data Model and associated data standards.
3. Use a modeling style that anticipates changes in RASSCLE II information and functional requirements.
4. Position the database to be useful for other purposes within the California Department of Health Services.
5. Strike a balance between extensibility, programming complexity, and system performance.

RIIDDM PHIN Compliance



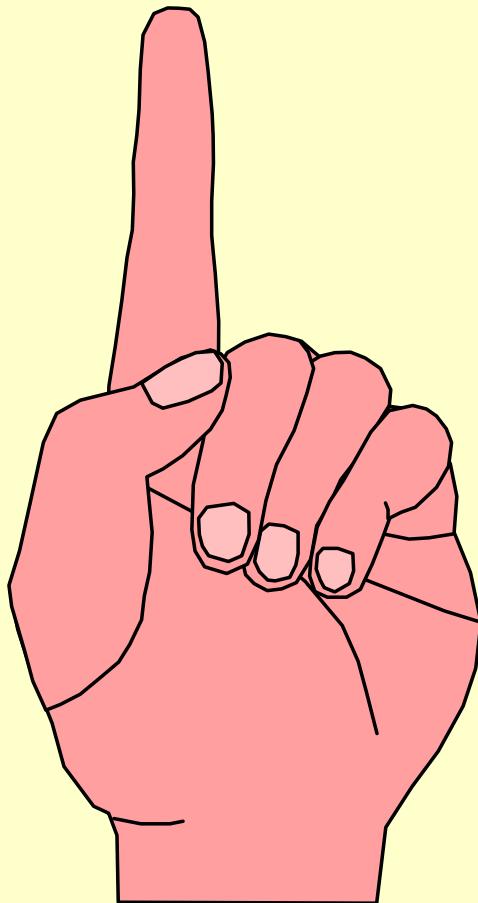
Compliance, with respect to the PHIN LDM, means that “derivative systems (the Database Data Models and Physical Data Models) are able to accurately represent each of the information concepts in their domains in a manner consistent with the PHIN LDM”. ~ *Public Health Information Network Logical Data Model Version 1.0; Center for Disease Control and Prevention; May 2004*

RIIDDM PHIN Compliance Guidelines

1. Each model element (table, column, constraint) must be traceable to the LDM or RIM. An exception to this would be model elements that are implementation artifacts not related to semantic information content.
2. Model elements must have consistent properties as their counterpart in the LDM or RIM. Properties include datatype, cardinality, vocabulary domain, coding strength, uniqueness, and null ability.
3. Business rules expressed in the LDM must be enforced by the PDM. The enforcement can be accomplished structurally or procedurally.
4. Structural differences between the LDM and the PDM that result in information loss should be assessed to ensure that no critical concepts have been inappropriately omitted.
5. The structures represented in the LDM should be able to be represented as views based on structures in the PDM.



Compliance Caveat



An important consideration to make when assessing the compliance of a PDM is that divergence from the PHIN LDM is not necessarily a bad thing. Divergence may be the result of requirements not previously considered. Such deviation should be reported to the CDC so that the PHIN LDM can be enhanced. Remember “All models are wrong. Some are useful”. Slavish adherence to the PHIN LDM is not recommended. However, careful consideration should be given to the pros and cons of any deviation. ~ *AbdulMalik Shakir, Principal Consultant, Shakir Consulting*

Acknowledgements

- Design of the R2 Database made possible through the support of the Lead Poisoning Prevention Branch at CDC.
- Final design of the R2 physical database will be implemented by Gensa Corporation, Shakir Consulting, and the CA-DHS/Childhood Lead Poisoning Prevention Branch.
- R2 System design and development will be completed by Northrop Grumman-IT, Gensa Corporation, and the CA-DHS/Childhood Lead Poisoning Prevention Branch.

Questions

